# Missouri's Regional COVID-19 Hospitalized Cases Model: Overview and Frequently Asked Questions

Model Overview

One of the many data analyses that inform Missouri's COVID-19 response is a regional model of hospitalized COVID-19 cases that the State of Missouri developed in partnership with the Washington University in St. Louis and Missouri Hospital Association.

Missouri's model uses a standard SEIR (susceptible, exposed, infectious, recovered) compartmental structure that is based upon a tool called LEMMA (Local Epidemiological Modeling for Management & Action), which was developed by experts from UMass Amherst, UC Berkeley, UCSF, and WUSTL. The model focuses on COVID-19 hospitalized cases to directly address the question of hospital capacity and provide a more accurate picture of COVID-19's impact on the healthcare system.
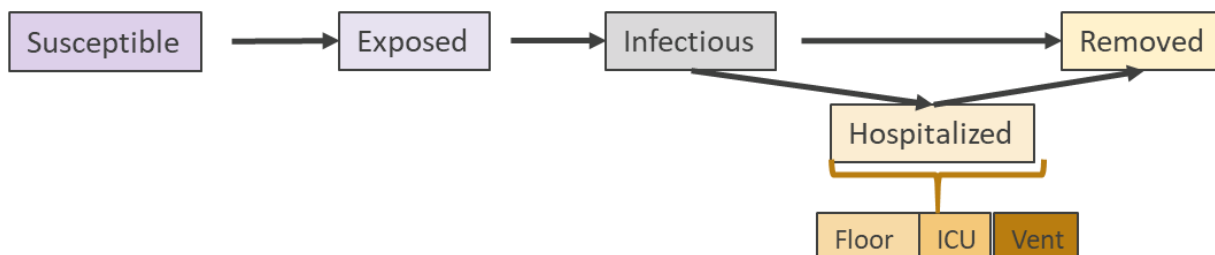
To help inform decisions at the regional and local level, each region is modeled separately using the latest local data, including COVID-19 confirmed and suspected hospitalizations, population, policy interventions, and average hospital length of stay.

General FAQs

Why are regional models of COVID-19 important?

When new diseases such as COVID-19 emerge, there is uncertainty about how best to control the epidemic. Decision makers must make their decisions with the best available information at hand.

Mathematical models are commonly used tools to help us understand how infectious diseases might impact key outcomes such as hospitalized cases or deaths. Today, there are many sophisticated models of COVID-19 that make global or national projections (e.g., see the Centers for Disease Control and Prevention's latest compilation of models at https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html). However, these generally do not incorporate key local or regional inputs, such as variations in local population demographics, healthcare system capacity, and social distancing and other policies. Regional projections are

important because regional epidemics may differ markedly from the national average and policy response occurs at state, county, and municipal levels.

Our model provides a range of plausible outcomes based on our current knowledge of COVID-19 in each region of Missouri with sufficient hospitalization and other relevant data. It also provides an approximate estimate of effective transmission rates across these regions. It can help inform decision making, as it portrays outcomes projected into the future based upon current assumptions, but it cannot actually predict what will happen in the future.

How were these geographical regions selected?

In Missouri, healthcare system data are typically analyzed by region because Missouri's hospitals serve patients across multiple counties.

The regions used for Missouri's COVID-19 model are based on Missouri's Highway Patrol Troop and Healthcare Coalition boundaries, which are historically used for healthcare preparedness and response planning. There are 2 notable differences:

1. Regions D, G, and I are combined into one Southwest region to reflect patient referral and EMS patterns, and their engagement with the Southwest Healthcare Coalition
2. While Perry County and Ste. Genevieve County reside in Highway Patrol Troop C (i.e., the Greater St. Louis Region), their data are reported through the Southeast Region due to their engagement with the Southeast Healthcare Coalition

A map of the counties in each region of analysis is included below.
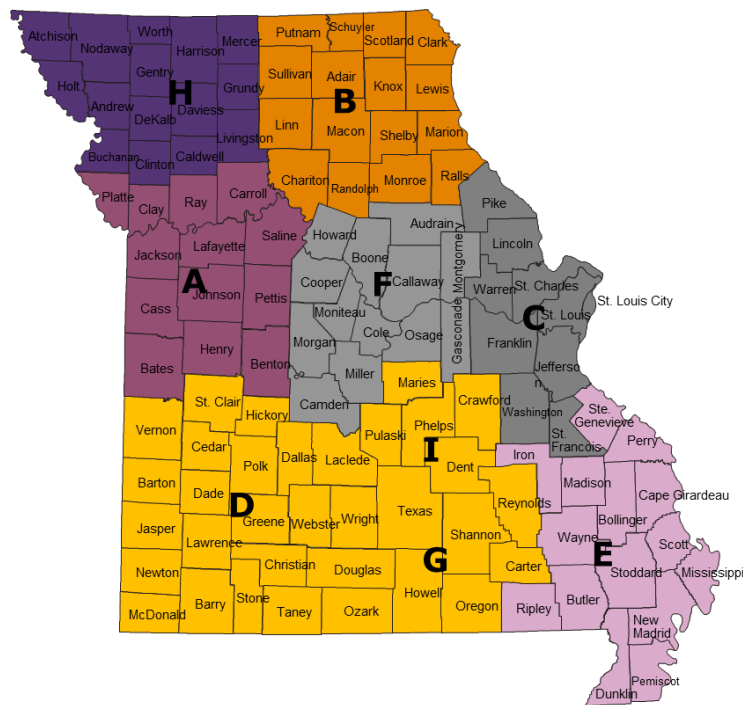
Why are no projections currently available for Regions B and H (Northwest and Northeast MO)?

Low levels of daily COVID-19 hospitalizations in the Northeast and Northwest regions limit the ability to model projections for these regions. In particular, the numbers of hospitalized cases have been so low that they have not been consistent with the SEIR model, which assumes at least some growth in cases over time. To date, hospitalized cases in these regions have been erratic, rising and falling sometimes but always with low overall numbers.

How far out does the model project hospitalized COVID-19 cases?

The model takes our current knowledge of COVID-19 to project hospitalized COVID-19 cases about a month into the future. The accuracy of projections decreases as we look further ahead, because even small deviations in the effective transmission rate will lead to diverging projections. This is because the SEIR model is nonlinear, with each infected case leading to new infections – and those new infections leading to additional new infections, and so on – according to the effective transmission rate. Restricting the projection window to a month balances the need for information to inform policy decisions with the importance of avoiding a false sense of precision.

How should you interpret these results?

These results provide a range of possible outcomes based on what we know right now, assuming that conditions remain the same, i.e. policies, people's behaviors, medical treatments, hospitalization rates, and the nature of the disease itself do not change. The model does not predict what will occur in the future, because any of those factors can potentially change in the weeks ahead. Therefore, the model is likely to be reasonably accurate for at least 1-2 weeks into the future. It may be less so when looking farther out into the future (see previous FAQ). Also, we should always be extra cautious about interpreting results in regions with low hospitalized case counts.

What do the model outputs represent?

The dots on the charts represent confirmed COVID-19 hospitalized cases in each region based on historical data. The model takes in the observed historical data on confirmed COVID-19 hospitalized cases (represented by the dots) to estimate future hospitalized cases. The black line represents the "best fit" scenario that plausibly explains the observed historical data. The grey dotted lines represent 25th and 75th percentile projected scenarios, and reflect the uncertainty behind model outputs.

Why are prior COVID-19 hospitalizations (dots) often not exactly on the "best fit" curve?

The "best fit" curve is the version of the SEIR model that is most consistent with the previously observed data. In other words, the SEIR model will always yield a smooth curve, but the exact data will never be completely consistent with that curve. At any given time, hospitalization rates (i.e. percent of cases requiring hospitalization) may fluctuate, as will the number of days each hospitalized case remains in the hospital. These factors are the main reasons why the model can't provide an exact fit of existing data.

Technical FAQs

What data goes into the model?

Demographic data on each region comes from the United States Census Bureau's 2014-2018 American Community Survey.

Hospitalized COVID-19 case data comes from 2 primary data sources:

1. From 3/14 to 4/24, confirmed COVID-19 hospitalized case counts by region came from Missouri Hospital Association (MHA)'s daily survey of Missouri hospitals
2. From 4/25 onward, we transitioned from MHA's survey to the Center for Disease Control's National Healthcare Safety Network (NHSN) COVID-19 Module. This shift in data source was done because the White House Coronavirus Task Force had requested that hospitals report similar data fields via NHSN, and we hoped to reduce the reporting burden on Missouri hospitals. That said, the data used is Missouri data as reported by Missouri hospitals.

How do the MHA and NHSN data on COVID-19 hospitalized cases differ?

While MHA's survey split out COVID-19 confirmed and suspected cases into 2 separate categories, NHSN reports a combined value for COVID-19 confirmed and suspected hospitalized cases.

Confirmed COVID-19 cases provides a better estimate of infection transmission and its impact on healthcare system capacity because many suspected COVID cases turn out to be negative and result in shorter lengths of stay. To approximate the likely share of COVID-19 confirmed cases from the total confirmed and suspected hospitalizations reported in NHSN, we analyzed how many suspected cases later turn positive and how quickly this happens to estimate a region-specific scaling factor. We then applied this scaling factor to the NHSN data.

From 6/1 onward, the NHSN data began to split out confirmed vs. suspected new COVID-19 hospital admissions. We are now starting to use this updated information into our model.

What statistical methods does the model employ?

The LEMMA model takes a Bayesian approach and is estimated with Monte Carlo computational methods.

Bayesian models in general specify a prior distribution for each parameter. This model allows the user to specify a range of possible options, where the most likely preliminary estimate has the most weight in the distribution. In particular, the most likely value receives 60% weight. A total of five values are specified with weights of 5%, 15%, 60%, 15%, and 5%.

A total of 1000 draws are made to the parameter space. This means 1000 draws from each parameter distribution, including disease parameters and policy parameters, are made. For each draw, an SEIR model projects what hospitalization counts would look like given those parameters. If the actual hospitalization data are consistent with the modeled hospitalizations, within a specified range, then the parameter draw is kept.

A posterior distribution is formed on the basis of all draws that are retained, i.e. all draws that could plausibly explain the observed data. The best fit is the median of all retained draws. Uncertainty is reflected in the 25th

percentile and 75th percentile bounds, which represent the range of projections consistent with retained parameters at the 25th and 75th percentiles in terms of fit with the observed data.

The posterior distribution of the reproductive rate provides an updated estimate of the effective reproductive rate and is used in combination with hospital length of stay to project hospitalized cases going forward, using the SEIR framework.

What are the model parameters?

The model contains many parameters that are estimated during each run and for each region that describe the nature of the illness. These include:

- initial reproductive rate of COVID-19
- days of pre-symptomatic infectiousness of an exposed individual
- length of infectiousness period
- time until hospitalization
- length of hospitalization if not in ICU
- percent of patients ventilated
- length of hospitalization if in ICU

Does the model account for policy interventions?

Yes, the model can detect changes in the effective reproductive rate that may ensue due to the implementation – or de-implementation – of local policies. The effect of the change is assumed to phase in over several days, where the number of days is also estimated within the model.

How does Missouri's COVID-19 model differ from other epidemiological models?

The model differs three main ways.

First, it is actually a set of models, because each region within Missouri is modeled separately, using its own total population as the initially susceptible population and using its own hospitalization data to fit a standard SEIR model. This has turned out to be a critical feature in understanding COVID-19 in Missouri, as our state has significant regional variation – especially due to differences in population density – that have resulted in different effective transmission rates. Moreover, different policy interventions have been implemented at different times in different regions, and the model is able to capture such variations.

The second difference is our reliance on hospitalized COVID-19 cases, as few other models have done. By focusing on hospitalized COVID-19 cases, we can directly answer questions on the healthcare system capacity (e.g., beds, PPE, and staff) needed to effectively manage the COVID epidemic in each region. This approach is not possible without a concerted effort at consistent, daily data collection, which may explain why few other models have used this type of data.

Finally, this model is dynamic in the sense of its ability to detect changes in COVID-19 over time that may be associated with changes in policies, people's behaviors, medical treatments, hospitalization rates, and the nature of the disease itself.